

# Data-Driven Power System Optimal Decision Making Strategy under Wildfire Events

Wanshi Hong  
Lawrence Berkeley  
National Laboratory  
[wanshihong@lbl.gov](mailto:wanshihong@lbl.gov)

Bin Wang  
Lawrence Berkeley  
National Laboratory  
[wangbin@lbl.gov](mailto:wangbin@lbl.gov)

Mengqi Yao  
University of  
California, Berkeley  
[mqyao@berkeley.edu](mailto:mqyao@berkeley.edu)

Duncan Callaway  
University of  
California, Berkeley  
[dcal@berkeley.edu](mailto:dcal@berkeley.edu)

Larry Dale  
University of  
California, Berkeley  
[lldale@lbl.gov](mailto:lldale@lbl.gov)

Can Huang  
Lawrence Livermore  
National Laboratory  
[huang38@llnl.gov](mailto:huang38@llnl.gov)

## Abstract

*Wildfire activities are increasing in the western United States in recent years, causing escalating threats to power systems. This paper developed an optimal and data-driven decision-making framework that improves power system resilience under wildfire risks. An optimal load shedding plan is formulated based on optimal power flow analysis. To avoid power system cascading failure caused by wildfire, we added additional transmission line flow constraints based on the identification of power lines with high ignition risk. Finally, a data-driven method is developed, leveraging multiple machine learning techniques, to model the complex correlations between input wildfire scenarios and the output power management strategy with significantly reduced computational complexities. The proposed data-driven decision-making framework can reduce the safety impacts on the electricity consumers, improve power system resilience under wildfire events.*

## 1. Introduction

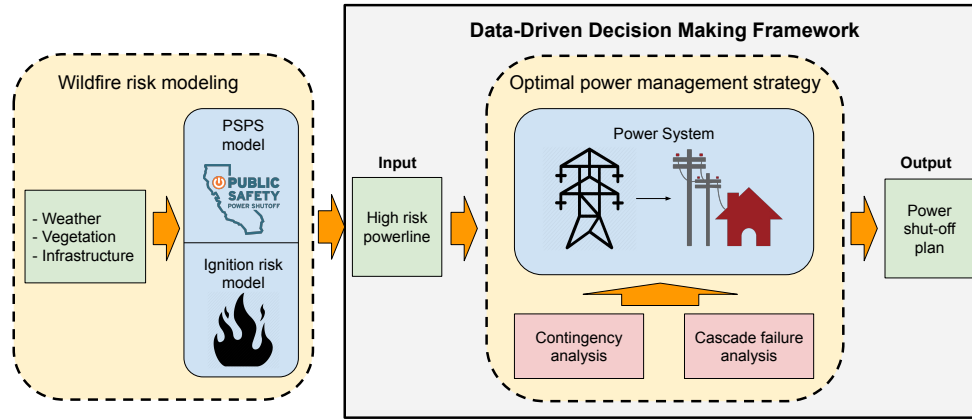
Wildfire events have caused escalating impacts all over the world in recent years. Wildfire-related disasters have increased in recent decades across the western United States due to drying weather, fuel accumulation over the past century and numerous ignition sources [1, 2]. By the end of 2020, 9,639 fires had burned 4,397,809 acres, making 2020 the largest wildfire season recorded in California's modern history.

Wildfire activities have a strong connection with power systems in the western United States as most power infrastructures are distributed in rural areas with high ignition risks [3]. Faults induced by grid assets are one of the major ignition risks that starts a wildfire.

The 2018 Camp Fire in California caused by powerline ignition killed 84 people and caused \$9.3 billion in housing damage, leading the file for bankruptcy by the responsible utility service holder Pacific Gas & Electricity (PG&E) [4]. Moreover, the spreading of a wildfire and the corresponding ignition prevention actions would also cause power outages, resulting in tremendous economic impacts. In October 2019, public power safety shut-off (PSPS) events turned off a million customer accounts in California during high ignition risk conditions to avoid wildfire events [5, 6]. While these PSPS events have reduced the wildfire risks, the economic and societal impact of these PSPS events is huge.

It is urgent to come up with optimal de-energization strategies for the power system under wildfire risks to improve public safety and reduce customer side impacts. This has become an emergent research topic in recent years with the increasing wildfire activities [7, 8]. While this is a relatively new research topic, we have summarized several state-of-the-art research works in this area.

There has been much research focusing on wildfire mitigation and wildfire risk analysis. In [3], the authors summarize the challenges and potential wildfire mitigation strategies in electric power grids. In [9], the authors presented a general framework to assess wildfire risk and mitigation options. They introduced the wildfire risk assessment process premised on several modeling approaches to characterize wildfire intensity and effects. In [10], the authors use the Markov Chain Monte Carlo sampling method to model the probability distribution of transmission line tripping under wildfire activities based on historic wildfire data. Wildfire monitoring is also a key issue in wildfire mitigation. In [11], the authors developed a tower location optimal selection



**Figure 1: Data-driven optimal power planning strategy framework.**

scheme for camera-based wildfire detection systems. In [12], the authors proposed a logic optimization method to improve the accuracy of satellite-based wildfire monitoring for power systems.

Wildfire-related power system management is a relatively new research topic that has drawn significant attention by researchers in recent years. There have been several attempts on power system de-energization, planning & management. In [13], the authors proposed a wildfire hazard prevention system based on big data mining techniques for power systems. In [14], the authors developed a power system planning strategy to reflect the impact of electric vehicle evacuations on power system resiliency during wildfire events. In [15], the authors modeled wildfire risks in power systems and developed an optimal power shut-off strategy using optimal power flow.

From the discussion above, it can be seen that the impact of PSPS events is significant. Moreover, the calculation of PSPS events for large grids is challenging and would take hours to solve optimal power flow problems. This motivates our work to develop an optimal and data-driven wildfire decision-making framework for the transmission system, which given the wildfire risk scenarios as inputs, and generate data-driven based optimal decision making strategy to improve power system resilience with significantly reduced computing time in real-time decision making. The detailed contributions include:

- Generate power system component (lines, substations, etc.) failure scenarios, given the ignition risk model;
- For different failure scenarios, perform power system analysis to generate ensembles of training data (input: wildfire risk distribution, grid

topology/connectivity, weather models, etc., output: optimal load shedding and power shut-off actions, etc;

- Design data-driven based decision-making framework based on optimal planning strategy developed.

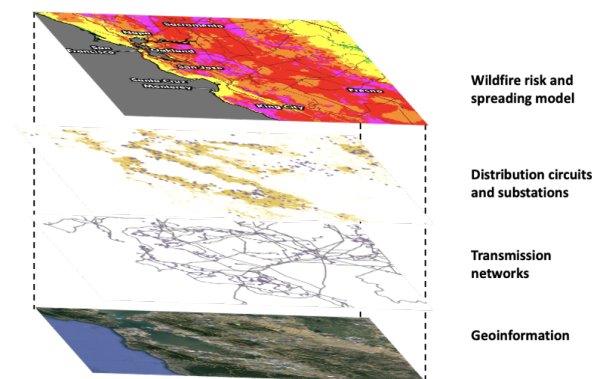
The rest of the paper is organized as follows: Section II introduces the problem formulation; Section III derives the optimal power management strategy; Section IV presents the data-driven based optimal power management design and shows the simulation results; Section V gives conclusions and discusses future works.

## 2. System Overview

As described in Section I, it is urgent to develop a fast and reliable decision making framework under wildfire scenario. In this paper, we propose to develop a data-driven based optimal power planning strategy under wildfire risk. This proposed method can be summarized in Fig. 1.

As shown in Fig. 1, we analyzed power system influence under wildfire ignition and public safety power shutoff (PSPS) events. Then, we developed an optimal power management strategy using the optimal power flow technique with cascading failure analysis considered as an additional constraint. Wildfire risk-based contingency analysis for this problem is studied. This optimal power management strategy is then built as an integrated module for the data-driven application. Based on historic wildfire data in California and wildfire risk impact on transmission lines, we then generated a list of input data that contains high wildfire risk scenarios and collect the output data which is calculated from the optimal decision based on the optimal power management framework. Finally, we use

data-driven tools to analyze the black box relationship between wildfire risk to power lines and optimal power management decisions. As a result, an overall framework between wildfire risk and optimal power management was studied and this study maps the effect of wildfire activity on power systems. Fig. 2 illustrates how this study maps wildfire risks to power systems. The objective of the wildfire risk model is to prevent wildfire caused by line ignition and list out potential higher risk lines to shut-off. For the data-driven decision-making model, the main objective is to reduce computational complexity, prevent cascading failure, and reduce the effect due to line shut-off. Note that line shut-off occurs in two scenarios: 1) high ignition risk; 2) effect of existing wildfire (that may not due to line ignition). The data-driven model will generate the best decision-making solution to reduce load shedding fastly instead of calculating the OPF problem for the entire network again (which may take a couple of hours). In summary, the overarching goal for this study is to develop data-driven optimal strategies that reduce calculation complexity and improve power system resiliency under wildfire, given the wildfire risks as inputs.



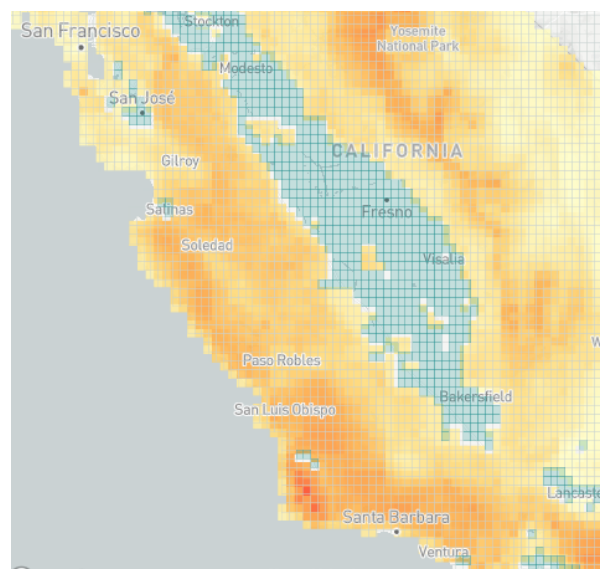
**Figure 2: Illustration of wildfire risk mapping to power systems.**

### 3. Optimal Load Shedding and Power Shut-off Formulation

In this section, the optimal power management strategy under wildfire risk is developed. First, the wildfire risk is modeled as the input to the optimal power management design. Then, an optimal power flow (OPF) problem is formulated to obtain the best power management strategy. Finally, a simulation study is presented to show the effectiveness of the proposed strategy.

### 3.1. Wildfire Risk Modeling

There exist two different causes that would lead to powerline interruptions during the wildfire season. The first is that the powerline causes fires and as a result, causing itself out of service. Under extreme weather conditions, transmission lines are prone to equipment failure and contact with vegetation. When the lines are energized and the surrounding environment is dry enough, ignitions triggered by powerlines could easily grow up to uncontrollable devastating wildfires. On the other hand, fires triggered by environmental or human causes would spread to powerlines under high wind conditions, resulting in energy interruptions.



**Figure 3: Illustration of a risk map generated using [16].**

To the best of our knowledge, the risk model for the powerline-induced wildfires has not been well developed yet. We propose to use data-driven methods to predict the probability of ignitions triggered by powerlines. In addition to weather and vegetation data sets, power grid infrastructure data (e.g., line length, voltage level, age of pole, size, and material of conduct) are collected to train the model. The ignition events in the power system happened rarely compared to non-ignition events, so how to deal with the imbalance data classification is one of the major challenges. On the other hand, how to allocate the data across temporal and spatial differentiation to achieve the best predicting results is a key research question. In our parallel work, we are exploring different machine learning models, for example, Random Forest and Neural Network, to identify the best model and imbalance strategy to achieve the highest classification score.

In this paper, we will use the risk maps generated by Westerling model [17]. The Westerling model is developed based on weather, vegetation, population, and historical fire data. One of the great features that the Westerling model has is that it can predict the fire probability considering the climate and population change. Fig. 3 shows an illustration of a fire risk map in the 2040s. Each grid cell has its probability, the darker the higher the fire probability would be. With the wildfire risk map obtained, we will identify the high risk areas in the power system and list out reasonable power shut off in the area as our power system input. After obtaining such a risk map, as illustrated in Fig 2, the next step will be mapping our transmission system to this risk map and as a result, we will obtain the fire risk for each powerline.

### 3.2. Optimal Power Flow Problem Formulation

**Objective Function.** The OPF problem can be described as minimizing load shedding at all buses subject to some constraints. The objective of the optimization problem is to reduce the amount of load shedding under wildfire risks, where power lines have been shut-off due to high ignition risk or close to an existing fire. The objective function is given as follows:

$$J = \sum_{t=1}^T \sum_{i=1}^n D_i(t), \quad (1)$$

where  $n$  is the number of bus,  $D_i(t)$  is the shedded load at bus  $i$  at time step  $t$ ,  $t = 1, \dots, T$ . The objective of the OPF problem is to minimize  $J$  to achieve the minimization of the load shedding during the wildfire risk time period  $[0, T]$ .

**Power flow equations.** The above OPF problem is built based on a power grid modeled by the triple  $\{\mathcal{N}, \mathcal{G}, \mathcal{L}, Y\}$  where  $\mathcal{N} = \{1, \dots, n\}$  is the set of buses,  $\mathcal{G} \subseteq \mathcal{N}$  is the set of generators with  $N$  generators inside,  $\mathcal{L} = \{1, \dots, L\}$  is the set of transmission lines. The status of each bus  $i$  is represented by its voltage  $V_i = |V_i|e^{j\theta_i}$ , where  $|V_i|$  is the voltage magnitude,  $\theta_i$  is the phase angle at bus  $i$ , and  $Y = G + jB \in \mathbb{C}^{n \times n}$  is the bus admittance matrix. The OPF problem is subjected to the power flow constraints:

$$P_i = \sum_{k \in \mathcal{N}} |V_i||V_k|(G_{ik} \cos(\theta_i - \theta_k) + B_{ik} \sin(\theta_i - \theta_k)) \quad (2)$$

$$Q_i = \sum_{k \in \mathcal{N}} |V_i||V_k|(G_{ik} \sin(\theta_i - \theta_k) - B_{ik} \cos(\theta_i - \theta_k)), \quad (3)$$

where we assume bus  $i = 1$  as the slack bus with  $\theta_1 = 0$ . For simplicity, we further assume there is only one generator at each bus, the apparent power at bus  $i$  is given by:

$$\begin{aligned} S_i &= P_i + jQ_i \\ &= \begin{cases} (P_i^g - P_i^d) + j(Q_i^g - Q_i^d) - D_i, & i \in \mathcal{G}, \\ -P_i^d - jQ_i^d - D_i, & i \notin \mathcal{G}, \end{cases} \end{aligned} \quad (4)$$

where  $P_i^g$  and  $Q_i^g$  are the generator active and reactive power at generator nodes,  $P_i^d$  and  $Q_i^d$  are the uncontrollable power sinks at all the buses.

**Cascading Failure Prevention Constraints.** To avoid cascading failure after shutting off the high-risk power components based on the given wildfire risk input, we added additional line capacity constraints given as follows

$$|P_j^l| \leq T_j^l, \forall j \in \mathcal{L} \quad (5)$$

where  $T_j^l$  is the power flow limit at line  $j$ ,  $P_j^l$  is the power flow at line  $j$ . The cascading failure constraint module is developed based on [18].

**Decision Variables.** The decision variable  $x$  for the OPF is:

$$x = [P_g^T, Q_g^T, D^T]^T, x \in \mathcal{X} \quad (6)$$

where  $P_g = [P_2^g, \dots, P_N^g]^T$  is the generator active power,  $Q_g = [Q_2^g, \dots, Q_N^g]^T$  is the generator reactive power,  $D = [D_1, \dots, D_n]^T$  load shed at each bus,  $\mathcal{X}$  is the bound for the decision variables including generator power limit and charger location limit.

To solve the OPF problem listed above, we used the genetic algorithm which is a heuristic approach to handle this non-linear and non-convex optimization problem [19].

### 3.3. Work Flow Description

The optimal power management work flow can be summarized in Algorithm 1.

### 3.4. Simulation Study

To test out the effectiveness of the proposed optimal power flow solution on a transmission network, we constructed the following simulation study on the

---

**Algorithm 1** Optimal Power Management Algorithm

---

```
1: Load wild fire risk input
2: Map wild fire risk to high risk powerlines
3: Remove high risk powerlines, obtain initial grid topology
4: Initialization  $i = 1$ , evaluate initial island condition
5: while island( $i$ ) exist do
6:   if island( $i$ ) is active then
7:     OPF analysis on island( $i$ ), obtain optimal power management plan ( optimal load shedding/generator operation)
8:   else
9:     Shut off components in island( $i$ )
10:  end if
11:   $i = i + 1$ 
12: end while
13: Output Optimal power management plan
```

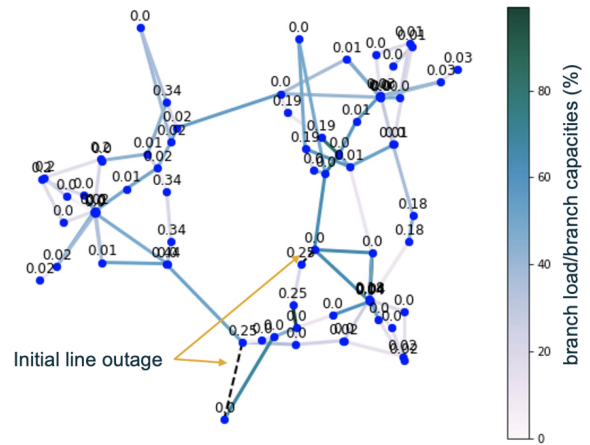
---

RTS-GMLC test case, which is a transmission system test case. In this simulation study, we compared the branch load status with and without the proposed OPF design. From which one can see that with optimal power management strategy with given high risk power line inputs, one can avoid line overload and prevent cascading failure.

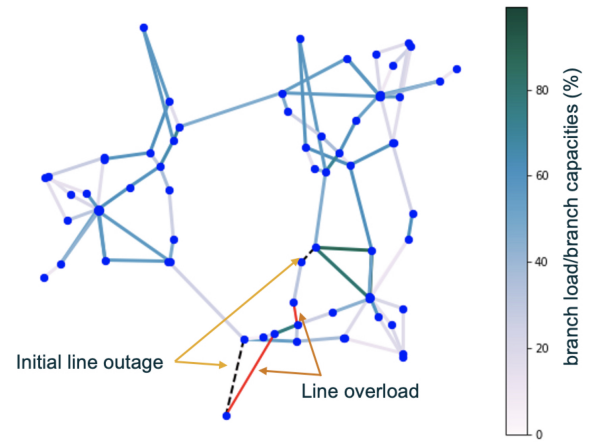
**Test on RTS-GMLC System.** We tested out our algorithm on the RTS-GMLC test case [20]. RTS-GMLC test case contains 73 nodes, 106 transmission lines, and 32 generators. This system can be divided into three connected regions, which are projected on a roughly 250\*250 mile region from Los Angeles to Las Vegas. A hydrologic time series load, wind, solar data is generated based on the reference region data collection. This test case is a good illustration of southwestern United States energy usage and weather condition.

**Simulation Results.** We first listed the powerlines with higher ignition risk and heavy load as potentially high-risk powerlines based on RTS-GMLC power system state data and randomly generated wildfire risk mapping. We assume one of the powerlines is under high risk (which is arbitrarily chosen from the list of high-risk powerlines) and needs power shut-off. We then apply our optimal power management strategy to perform load shedding. Fig. 4 and Fig. 5 shows our optimal power management results.

Fig. 4 shows the line overload prevention behavior. The initial wire down is shown in the dashed line. From which we can observe that with optimal power management in Fig. 4a, the power system will perform load shedding (with the amount shown above each



(a) With optimal power management.



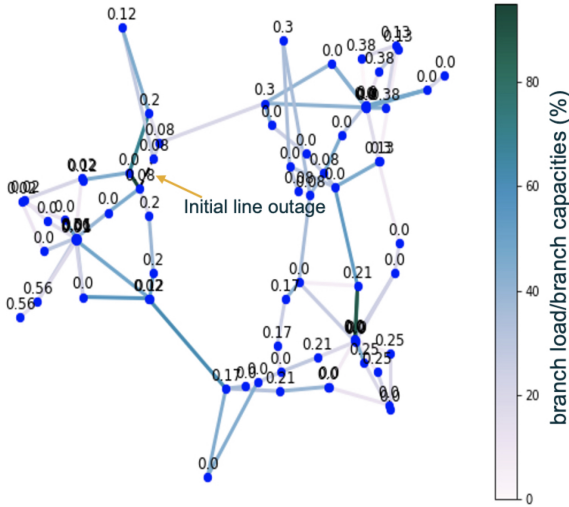
(b) Without optimal power management.

**Figure 4: Line overload prevention.**

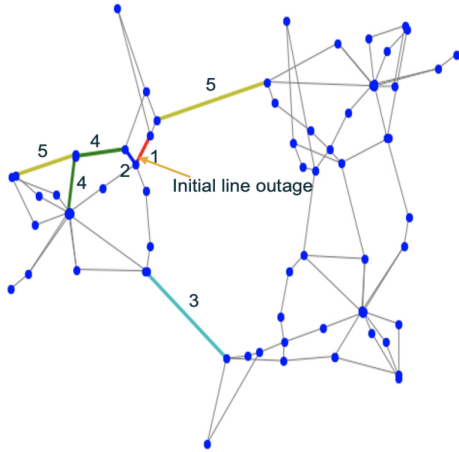
node). The powerlines in the network will not overload as all branch loading is under maximum line capacity. However, under the same power system operation condition, without optimal power management shown in Fig. 4b the red lines will overload, leading to cascading failure in the power system and potentially cause blackouts. This result shows that with the proposed optimal power management strategy line overload can be prevented.

Fig. 5 shows the cascading failure prevention behavior. The initial line wire down is shown in dash line in Fig.5a and the same line is marked in red as the first stage in cascading failure analysis in Fig. 5b. From which we can observe that with optimal power management in Fig. 5a, the power system load shedding amount is shown above each node. The powerlines will not overload. Under the same power system operation condition, cascading failure will occur without optimal power management. The cascading stages are marked





(a) With optimal power management.



(b) Without optimal power management.

Figure 5: Cascading failure prevention.

in Fig. 5b shown in red, blue, azure, green, and yellow, respectively. Eventually, the power system will break out into three islands. This result shows that the proposed scheme can prevent power system cascading failure.

### 3.5. Summary

In summary, this section presented the optimal power management design, where we considered wildfire risk as input and formulate optimal power flow problems to reduce customer influence and prevent power system cascading failure. This algorithm is then tested on the RTS-GMLC test case. The simulation results show the effectiveness of the proposed algorithm. In the real-world implementation in power systems,

the above algorithm can be directly used for power system management. However, when monitoring a large network under wildfire risk, the real-time contingency analysis is time-consuming. To handle this problem, we propose to use a pre-trained data-driven decision making framework. In the next section, this proposed optimal power management strategy serves as the black box decision-making framework to be learned by the data-driven approach.

## 4. Data-Driven based Optimal Decision Making Framework Formulation

In this section, the data-driven based optimal decision-making framework is formulated. First, the machine learning problem is defined with system input/output introduced and the data collection scheme designed. Then, the classification methodology is introduced. Finally, a simulation study is presented to show the effectiveness of the proposed scheme.

### 4.1. Data-Driven Problem Modeling

To model the optimal power management strategy into a data-driven problem, dataset acquisition is needed. First, we will choose the input features and output labels for the problem. Then we will introduce the data collection scheme based on risk-based contingency analysis. Finally, we will state the objective of the problem. This problem is tested and the dataset is collected based on the RTS-GMLC test case. The data-driven problem model is shown in Fig. 6.

**Input features.** The input features for the problem include generator status ( $P_g$ ) at each bus, load profile ( $P_d$ ,  $Q_d$ ) at each bus, and powerline ignition risk topology (which shows the connectivity of the grid network with high wildfire risk lines removed).

The powerline ignition risk topology is obtained from the wildfire risk model described in Section 3.2. We mapped the wildfire risk from the Westerling model to the transmission system to build the high ignition risk powerline set. For each input, we randomly select one, two, or three lines from this set to be down due to high ignition risk and generate the powerline ignition risk topology.

**Label selection.** To simplify the problem and improve training performance on a limited amount of data, we formulated the optimal power management strategy into a classification problem with some threshold chosen for optimal decision making. Moreover, to balance the dataset, we developed a two-level labeling structure. We first consider cascading failure classification to reduce the number of cases that

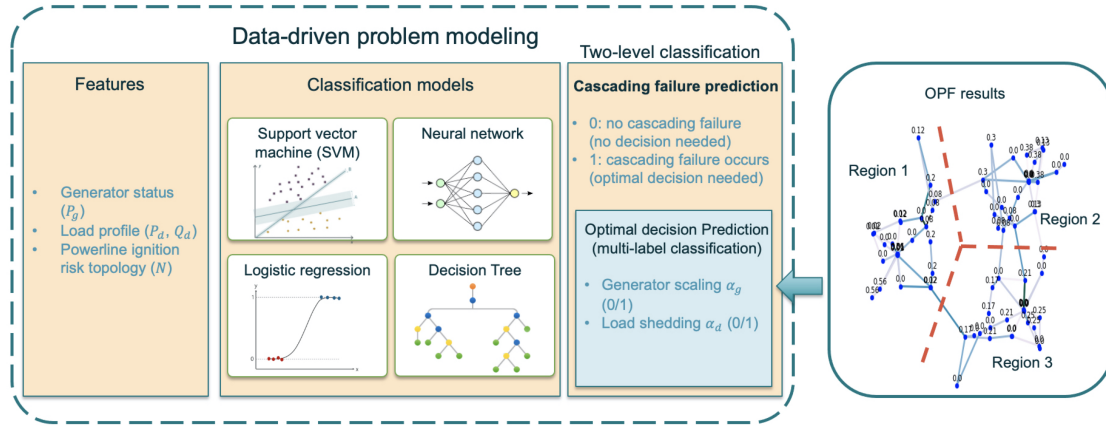


Figure 6: Data-driven problem formulation.

will not trigger cascading failure, where no optimal decision is needed. Then for those cases that will trigger cascading failure, we perform a multi-label-based optimal decision classification.

**Level 1: Cascading failure prediction.** The first-level classification problem is for cascading failure prediction. We test out a list of input scenario cases with high wildfire risk powerlines down to see whether cascading failure will occur. The output label is 0/1 (0: no cascading failure, 1: cascading failure occurs).

**Level 2: Optimal decision prediction.** The second-level classification problem is for optimal decision prediction. We first divided the RTS-GMLC model into three regions. In each region, the decision is made based on the optimal power management results. If the maximum generator power variation in region  $i$  exceed  $\alpha_g$ , then the generator scaling label for region  $i$  is True (or 1),  $i = 1, 2, 3$ . Similarly, if the maximum load shedding amount in the region  $i$  exceeds  $\alpha_d$ , then the load shedding label for region  $i$  is True (or 1), and vice versa.  $\alpha_g, \alpha_d \in (0, 1)$  are threshold parameters. In real-world implementation, this optimal decision prediction can provide real-time power management decisions based on given wildfire risk inputs. With the optimal decision predictions implemented to the power system, the system operation condition will update and the new branch load will rearrange within each line limitations to avoid line overload as shown in Fig. 4.

**Data Collection and Contingency Analysis.** We first listed out the powerlines under high wildfire risks and randomly downs one, two, or three powerlines from the list. This mimics initial ignition around certain lines that cause wire down event. We collected  $M$  set of wire down cases. Then, we randomly generate  $M$  set of power system states (including generator power and load profile) within the range of operation for the  $M$  wire

down cases. This mimics random energy usage behavior at any time of the day. This formulates the input dataset for the study. Then, we perform risk-based contingency analysis for the above  $M$  test cases using the optimal power management strategy developed in Section 3. For each test case, the output label is chosen based on the criteria above. As a result, we will obtain  $M$  set of data. For the RTS-GMLC test case, there are 238 input features (including line connectivity, generator status, and load profiles), 6 output features for optimal decision prediction, and 1 output feature for cascading failure prediction. We generated 37,000 sets of data for cascading failure prediction and 2,700 sets of data for optimal decision prediction.

**Computational Complexity Discussion.** Using OPF formulation to make an optimal decision in real-time has a time complexity of the order  $O(n^2)$ , where  $n$  is the number of buses in the transmission network. For conventional  $n$ -k contingency analysis, the time complexity is  $O(\binom{n}{k})$ . Compared with the proposed data-driven approach, the time complexity for data collection and training is  $O((M + 1) * n^2)$ . However, the time complexity to make one decision in real-time with a pre-trained model is  $O(1)$ . As a result, the proposed data-driven decision-making framework can significantly reduce real-time decision-making computation complexity compared with conventional power system decision-making strategies.

## 4.2. Methodology

To improve training performance, we tested out several machine learning models for this multi-label classification problem.

**Support vector machine (SVM).** Support vector machine is a supervised learning method that shows

significant advantage in high dimensional spaces [21, 22]. SVM is one of the most robust prediction methods being developed based on statistical learning frameworks proposed by Vapnik and Chervonenkis. For this study, the input feature dimension is high and the amount of data is limited. SVM seems to be a proper solution to this problem. This implementation is based on LIBSVM [23], with linear kernel chosen and regularization parameter  $C = 1$ .

**Neural Network (NN).** The neural network is a popular machine learning tool that is inspired by the biological neural networks that mimic animal brains [24, 25]. Multi-layer perceptrons (MLP) is one of the most commonly used neural networks and it is capable of nonlinear function approximation. MLP is built by layers of neurons which consists of nonlinear activation functions. MLP is capable to learn non-linear models which suit our problem. In this implementation, we used a multi-layer perceptron classifier in sklearn toolbox [26] in python using LBFGS solver for optimization with two hidden layers (100 and 50 neurons, respectively)

**Logistic Regression.** Logistic regression is a linear classification model [27, 28]. Logistic regression can provide probability predictions and it requires less training efforts, making it suitable for larger datasets. Logistic regression has a great potential advantage as we move to bigger power system test cases in the future. In this implementation, we used logistic regression classifiers with SGD training and L2 regulation in sklearn toolbox [26, 29].

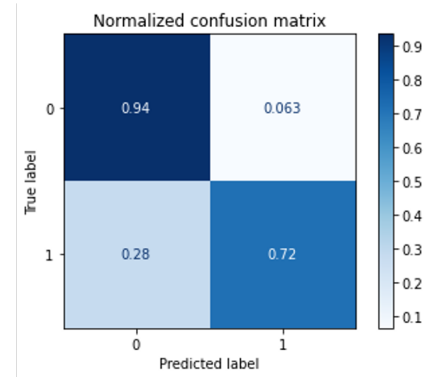
**Decision Tree.** The decision tree is a non-parametric supervised learning method that predicts targets by learning decision rules based on data features [30, 31]. The decision tree is one of the most popular machine learning methods as the implementation logic is simple and it serves as a baseline method for this study. In this implementation, we used a decision tree classifier from sklearn toolbox [26] with supported criteria being Gini impurity and using best split strategy.

#### 4.3. Simulation Study

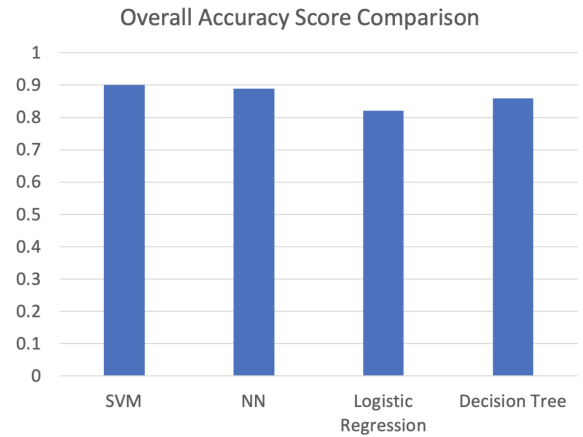
In this section, simulation results are presents to show the performance of the proposed strategy. We tested out the data-driven algorithm on the RTS-GMLC test case. We generated 37,000 sets of data for cascading failure prediction and 2,700 sets of data for optimal decision prediction. The results are shown below.

**Cascading failure prediction.** Fig. 7 shows cascading failure prediction results. Fig. 7a shows the normalized confusion matrix using support vector machine and Fig. 7b shows the overall accuracy score

comparison among different methods. From which one can see that an overall accuracy between 80% - 90% is achieved using different training methods. Support vector machine achieves a best overall accuracy score of 90%. The definition of overall accuracy is the total number of correct predictions divided by the total number of cases (both cascade failure and non-cascade failure). The confusion matrix shows the prediction accuracy among different classes and SVM can achieve 94% accuracy for non-cascading failure event prediction.



(a) Normalized confusion matrix for SVM.



(b) Overall accuracy score comparison.

Figure 7: Cascading failure prediction.

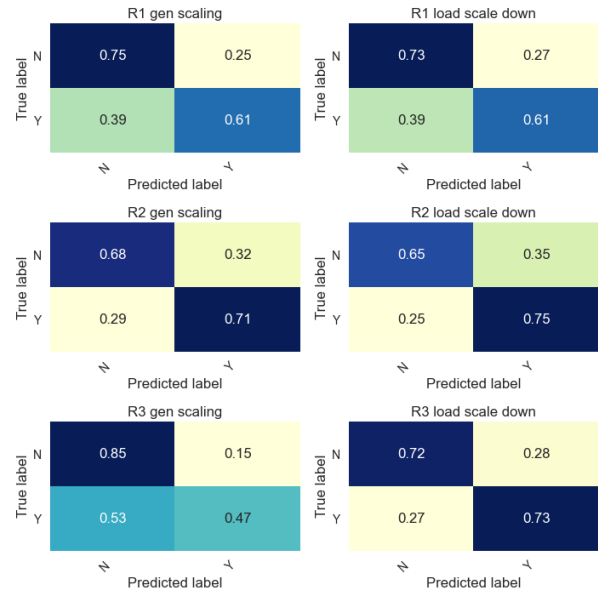
**Optimal decision prediction.** For optimal decision prediction we choose the threshold for generator scaling and load shedding as  $\alpha_p = 10\%$  and  $\alpha_d = 5\%$ . The choice of threshold is not only used to reflect the optimal decision but also used to balance the dataset. Fig. 8 shows cascading failure prediction results. Fig. 8a shows the normalized confusion matrix using SVM and Fig. 8b shows the overall accuracy score comparison among different methods. From which one can see that an overall accuracy between 50% - 70% is achieved



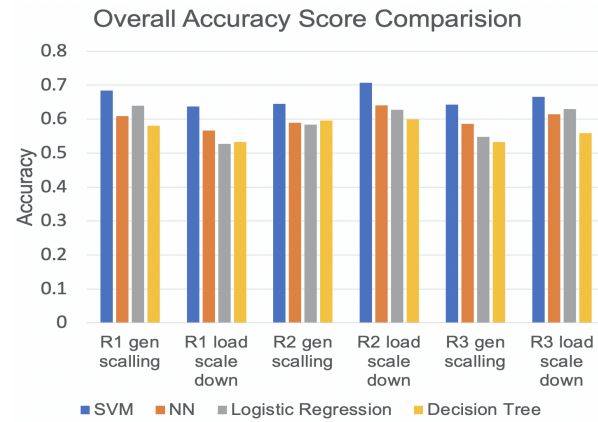
for each label using different training methods. The overall accuracy for each label is defined as the total number of correct predictions of each class (an example of one label is: region one load scale down by 5% is true) divided by total predictions of this class. Support vector machine achieves the best overall accuracy score for all the labels. The confusion matrix for different labels shows the accuracy by class and SVM can achieve 85% accuracy for generator scaling prediction in region 3. The accuracy for optimal decision is not as good as cascading failure prediction as this problem is a multi-label classification problem with all the labels having inter-dependency. Moreover, to achieve a better training performance, more data is needed as cascading failure prediction has 13 times more data than optimal decision prediction. For future work, we will use high-performance computing tools to expedite the data generation process.

## 5. Conclusion and Future Work

This paper developed a data-driven based power system optimal decision-making strategy under wildfire activities. First, the problem is formulated and the framework is introduced. Then the optimal decision-making strategy used for data generation is designed and tested. Finally, the data-driven decision-making strategy is developed and tested on the RTS-GMLC test case. We achieved 90% accuracy on cascading failure prediction and an average of 67% overall accuracy for all labels on optimal decision prediction using SVM. This data-driven decision-making framework can reduce the safety impacts on the electricity consumers, improve power system resilience under wildfire events and reduce the computational complexity for real-time power shut-off decision making. Simulation results show the effectiveness of the proposed data-driven decision-making framework. To improve the performance of the proposed scheme, some potential future work includes collecting more data using high-performance computing tools and using the deep graph learning method to learn the connectivity features of the power grid. In the current study, static cascading failure analysis based on studies from [18] is used instead of transient cascading failure analysis as static analysis is more straightforward to collect data for data-driven framework development. It is a good future work to consider a more sophisticated cascading failure analysis approach. Moreover, we plan to add a wildfire spreading model with a dynamic decision making procedure to more accurately identify the power system operation decision under fire spreading events



(a) Normalized confusion matrix for SVM.



(b) Overall accuracy score comparison.

Figure 8: Optimal decision prediction.

In the next phase of this work, we plan to test out the algorithm in larger networks such as ACTIVSg10k test case [32] using high-performance computing tools.

## 6. Acknowledgement

This research is supported by the UC-Lab fee program, Lawrence Berkeley National Laboratory, and Lawrence Livermore National Laboratory.

## References

- [1] J. T. Abatzoglou, C. M. Smith, D. L. Swain, T. Ptak, and C. A. Kolden, "Population exposure to pre-emptive de-energization aimed at averting wildfires in northern california," *Environmental Research Letters*, vol. 15,

- no. 9, p. 094046, 2020.
- [2] V. C. Radeloff, D. P. Helmers, H. A. Kramer, M. H. Mockrin, P. M. Alexandre, A. Bar-Massada, V. Butsic, T. J. Hawbaker, S. Martinuzzi, A. D. Syphard, *et al.*, "Rapid growth of the us wildland-urban interface raises wildfire risk," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. 3314–3319, 2018.
  - [3] J. W. Muhs, M. Parvania, and M. Shahidehpour, "Wildfire risk mitigation: A paradigm shift in power systems planning and operation," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 366–375, 2020.
  - [4] D. M. F. C. T. Jeffery, S. Yerkes and R. Turakhia, "2019 wildfire risk report," tech. rep., 2019. Available: <https://www.corelogic.com/insights-download/wildfire-risk-report.aspx>.
  - [5] P. Gas and E. Company, "Pacific gas and electric company amended 2019 wildfire safety plan," tech. rep., 2019.
  - [6] S. C. E. Company, "Southern california edison company's 2019 wildfire mitigation plan," tech. rep., 2019.
  - [7] S. Jazebi, F. De Leon, and A. Nelson, "Review of wildfire management techniques—part i: Causes, prevention, detection, suppression, and data analytics," *IEEE Transactions on Power Delivery*, vol. 35, no. 1, pp. 430–439, 2019.
  - [8] S. Jazebi, F. De Leon, and A. Nelson, "Review of wildfire management techniques—part ii: Urgent call for investment in research and development of preventative solutions," *IEEE Transactions on Power Delivery*, vol. 35, no. 1, pp. 440–450, 2019.
  - [9] J. H. Scott, M. P. Thompson, and D. E. Calkin, "A wildfire risk assessment framework for land and resource management," 2013.
  - [10] J. Lu, J. Guo, Z. Jian, and X. Xu, "Research and application of efficient risk analysis method for electric power grid multiple faults under widespread wildfire disasters," *International Transactions on Electrical Energy Systems*, vol. 29, no. 9, p. e12055, 2019.
  - [11] A. Heyns, W. Du Plessis, M. Kosch, and G. Hough, "Optimisation of tower site locations for camera-based wildfire detection systems," *International journal of wildland fire*, vol. 28, no. 9, pp. 651–665, 2019.
  - [12] Z. Zhou and G. Chen, "Research on logic optimization and reliable calculation model of satellite based wildfire monitoring for power transmission line," in *IOP Conference Series: Earth and Environmental Science*, vol. 192, p. 012019, IOP Publishing, 2018.
  - [13] T. Zhou, B. Li, C. Wu, Y. Tan, L. Mao, and W. Wu, "Studies on big data mining techniques in wildfire prevention for power system," in *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2)*, pp. 866–871, IEEE, 2019.
  - [14] D. L. Donaldson, M. S. Alvarez-Alvarado, and D. Jayaweera, "Power system resiliency during wildfires under increasing penetration of electric vehicles," in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6, IEEE, 2020.
  - [15] N. Rhodes, L. Ntamo, and L. Roald, "Balancing wildfire risk and power outages through optimized power shut-offs," *IEEE Transactions on Power Systems*, 2020.
  - [16] "Cal-Adapt." <https://cal-adapt.org/tools/wildfire/>, 2018.
  - [17] A. L. Westerling, *Wildfire Simulations for California's Fourth Climate Change Assessment: Projecting Changes in Extreme Wildfire Events with a Warming Climate: a Report for California's Fourth Climate Change Assessment*. California Energy Commission Sacramento, CA, 2018.
  - [18] H. Cetinay, S. Soltan, F. A. Kuipers, G. Zussman, and P. Van Mieghem, "Comparing the effects of failures in power grids under the ac and dc power flow models," *IEEE Transactions on Network Science and Engineering*, vol. 5, no. 4, pp. 301–312, 2017.
  - [19] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," 1988.
  - [20] C. Barrows, A. Bloom, A. Ehlen, J. Ikäheimo, J. Jorgenson, D. Krishnamurthy, J. Lau, B. McBenett, M. O'Connell, E. Preston, *et al.*, "The ieee reliability test system: A proposed 2019 update," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 119–127, 2019.
  - [21] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152, 1992.
  - [22] W. S. Noble, "What is a support vector machine?," *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
  - [23] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.
  - [24] W. S. Sarle, "Neural networks and statistical models," 1994.
  - [25] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric environment*, vol. 32, no. 14–15, pp. 2627–2636, 1998.
  - [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
  - [27] R. E. Wright, "Logistic regression," 1995.
  - [28] S. Sperandei, "Understanding logistic regression analysis," *Biochemia medica*, vol. 24, no. 1, pp. 12–18, 2014.
  - [29] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *the Journal of machine Learning research*, vol. 9, pp. 1871–1874, 2008.
  - [30] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
  - [31] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275–285, 2004.
  - [32] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on power systems*, vol. 32, no. 4, pp. 3258–3265, 2016.